

Evolving Concepts of Bacterial Species

Timothy G. Barraclough · Kevin J. Balbi ·
Richard J. Ellis

Received: 3 December 2011 / Accepted: 5 April 2012 / Published online: 22 April 2012
© Springer Science+Business Media, LLC 2012

Abstract The same evolutionary forces that cause diversification in sexual eukaryotes are expected to cause diversification in bacteria. However, in bacteria, the wider variety of mechanisms for gene exchange (or lack thereof) increases the range of expected diversity patterns compared to those of sexual organisms. Two parallel concepts for bacterial speciation have developed, based on ecological divergence or barriers to recombination in turn. Recent evidence from DNA sequence data shows that both processes can generate independently evolving groups that are equivalent to sexual species and that represent separate arenas within which recombination (when it occurs), selection and drift occur. It remains unclear, however, how often different processes act in concert to generate simple units of diversity, or whether a more complex model of diversity is required, specifying hierarchical levels at which different cohesive processes operate. We advocate an integrative approach that evaluates the effects of multiple evolutionary forces on diversity patterns. There is also great potential for laboratory studies of bacterial evolution that test evolutionary mechanisms inferred from population genetic analyses of multi-locus and genome sequence data.

Keywords Bacteria · Prokaryotes · Species · Speciation · Diversification · Recombination · Divergent selection · Horizontal gene transfer · Lateral gene transfer · Multi-locus sequence types · Genomes · Niches

T. G. Barraclough (✉) · K. J. Balbi
Department of Life Sciences, Imperial College London,
Silwood Park Campus, Ascot, Berkshire SL5 7PY, UK
e-mail: t.barraclough@imperial.ac.uk

R. J. Ellis
Animal Health and Veterinary Laboratories Agency, AHVLA
Weybridge, New Haw Addlestone, Surrey KT15 3NB, UK

Introduction

DNA sequence data are increasingly used to characterise biodiversity, not least in groups of organisms in which traditional taxonomic approaches have proved of limited use (Tautz et al. 2003). Bacteria (and other prokaryotes) are particularly dependent on DNA approaches and particularly challenging for systematics. DNA sequencing of environmental samples has revealed that only a tiny fraction of bacteria are culturable and amenable to traditional taxonomic methods (Rappe and Giovannoni 2003). Moreover, extrapolation from available samples indicates that the true global species richness of bacteria, as defined by current methodology, could number in the billions (Curtis et al. 2006). However, although a wealth of sequence data for bacteria is becoming available, there remain major challenges in characterising bacterial diversity and understanding its evolutionary causes and consequences.

- (1) Bacteria offer conceptual challenges to theories of species and speciation, because of major differences in their mode of inheritance from typical multicellular eukaryotes (Gevers et al. 2005). For example, bacteria are clonal, yet many of them exchange DNA either by homologous recombination, which occurs most often between closer relatives, or by horizontal gene transfer, which can transfer ecologically important genes even between distantly related strains via plasmids or transposons (Ochman et al. 2005; Norman et al. 2009). This complexity creates difficulties in applying concepts of species developed for plants and animals (Coyne and Orr 2004). Nonetheless, the same basic processes can cause diversification in bacteria as in multicellular eukaryotes (Cohan 2001): geographic isolation (Whitaker 2006), reproductive isolation

(i.e. genetic mechanisms preventing exchange of DNA, at least for core genes, Fraser et al. 2007), and divergent selection (i.e. ecological speciation, Cohan 2001). A key question in bacteria is to what extent do these mechanisms act in concert on components of the genome, to produce diversity units equivalent to eukaryotic species, rather than acting separately on different genes.

- (2) Most broad-scale environmental surveys have used a phenetic approach to delimit species, i.e. groups of individuals separated by sequence divergence above a certain threshold, now using 1 % for 16S rRNA (Schloss and Handelsman 2005; Stackebrandt and Ebers 2006) or 6 % for protein-coding genes (Venter et al. 2004). This assumes that a single threshold can be used across all bacteria, for which there is little theoretical or empirical justification. A more direct approach is to test for signatures of independent evolution in patterns of DNA variation and to use these signatures to infer units of diversity. A few studies recognised this potential (Acinas et al. 2004; Whitaker and Banfield 2006), but still relied on threshold approaches to delimit operational taxonomic units (OTUs). An alternative approach is to use a statistical framework for hypothesis testing and estimation of the biodiversity of a sample (Koeppel et al. 2008; Barraclough et al. 2009). Equivalent methods for optimising evolutionarily significant taxa are needed for multi-locus and whole genome data as well.
- (3) Most broad environmental surveys have relied on single genes, notably 16S rRNA (Acinas et al. 2004). Strengths of this approach are that the sequences are easy to place within the known phylogeny of bacteria. Weaknesses include reliance on a single gene that might not reflect more complex patterns of inheritance (Eisen 2007). The first multi-locus approaches were developed for pathogenic bacteria (Hanage et al. 2006). Multi-locus sequence typing is feasible for bacteria that can be isolated as individuals from the environment, i.e. by culturing or isolating single cells (Papke et al. 2007), and increasingly this will incorporate whole genome sequencing. The cutting edge is to sequence metagenomes from whole communities, which provides a vast treasure-trove of information on functionally interesting genes (Qin et al. 2010). However, it is difficult from such data to piece together individual genomes sufficiently for evolutionary methods of species delimitation (although future developments might allow individual genomes to be sequenced even from unculturable isolates).

Here, we discuss theoretical concepts of diversification in bacteria and how they might be tested using DNA

sequence data. Two parallel evolutionary concepts of species in bacteria have developed, one focused on ecological divergence and the other on barriers to recombination. We discuss how these alternative mechanisms interact in bacteria with varying levels and types of recombination, and how their relative importance can be tested. We argue that progress in understanding evolutionary mechanisms generating discrete clustering in bacteria requires an integrated approach evaluating both ecological divergence and recombination. Our intention is to summarize the current state of the field and to offer ideas for future progress, which we hope will stimulate microbial researchers and evolutionary biologists alike: for comprehensive recent reviews of bacterial speciation see Achtman and Wagner (2008), Doolittle and Zhaxybayeva (2009) and Wiedenbeck and Cohan (2011).

General Concepts

Speciation concerns the splitting of a single gene pool into two. In sexual organisms, the focus is on interbreeding: interbreeding maintains cohesion within species but reproductive isolation allows divergence between them (Coyne and Orr 2004). Sometimes selection might be strong enough to allow divergence even with interbreeding. In this case, divergence might be restricted to genomic regions linked to genes responsible for, for example, ecological differences among species (Nosil et al. 2009), unless ecological divergence itself leads as a by-product to reduced interbreeding. Via either reproductive isolation or divergent selection, species represent separate arenas for recombination, selection and drift and therefore evolve independently: mutations arising within one of the species do not spread into the second species. In strictly clonal organisms, recombination never occurs, but selection and drift maintain cohesion within populations and cause divergence between them (Cohan 2001; Barraclough et al. 2003). If barriers arise that prevent a genotype spreading and replacing individuals in a separate population, then this creates conditions for divergence and the emergence of distinct genetic and phenotypic clusters. This has been referred to as demographic (or ecological) non-exchangeability (Templeton 1989) or independent limitation (Barraclough et al. 2003). In this scenario, species constitute separate arenas for selection and drift (Fisher 1930).

Barriers to Recombination

Christophe Fraser, William Hanage and Brian Spratt developed neutral models of speciation in bacteria with varying rates of recombination (Fraser et al. 2007). These

models considered a scenario in which recombination occurs randomly throughout the genome through homologous recombination via transformation (Lorenz and Wackernagel 1994; Vos 2009). All isolates were considered to be ecologically equivalent. When the recombination rate is lower than the mutation rate per gene, bacterial populations are effectively clonal. When recombination equals or exceeds the mutation rate, bacterial populations behave as ‘sexual’ organisms with the result that recombination acts to shuffle gene combinations and prevents individuals within a population from become too genetically divergent from one another (Doroghazi and Buckley 2010; Keymer and Boehm 2011). Distinct species evolve only if barriers to gene exchange arise, equivalent to reproductive isolation in sexual eukaryotes, or if divergent selection is strong enough to counteract the homogenizing force of recombination.

This model assumes that there is no divergent selection on different genotypes and, as with all ecological neutral models, might not be well supported in natural settings (for example, typical genetic diversity in bacterial populations seem too low to be explicable by neutral processes alone given their large effective population sizes). However, it provides a null model against which to judge the action of selection. Furthermore, bacterial communities typically display a pattern of genetic clustering even when surveyed with neutral genetic markers such as house-keeping genes that do not appear to be under divergent selection between clusters (e.g. Acinas et al. 2004). For clustering to be observed at such loci in taxa with levels of recombination above the thresholds indicated by the neutral models of Fraser et al. (2007), requires either that those marker genes are closely linked to genes under selection, or that mechanisms are present that reduce rates of recombination between distinct clusters.

One mechanism for reducing gene exchange between distinct clusters is that the frequency of homologous recombination declines with increasing genetic distance (Roberts and Cohan 1993; Zawadzki et al. 1995). In this scenario, non-recombining groups might emerge spontaneously within a population if isolates can diverge enough genetically for a reduction in recombination rates to occur. This mechanism alone, however, would not be enough to explain long-term coexistence of non-recombining groups. First, it requires that mutation rates are high enough, and that the decline in recombination with genetic divergence is steep enough, that gene regions can accumulate sufficient genetic divergence to reduce recombination rates, while at the same time recombination is acting to reduce divergence. Empirical evidence indicates that this assumption is rarely met (Fraser et al. 2007). Second, even if these conditions were met, if non-recombining groups still occupied the same ecological niche, then descendants of

one of the groups should displace descendants of the other group by chance (i.e. drift occurring within the wider population). Processes of mutation within the wider population might throw out non-recombining groups at a certain rate, but there would be ongoing turnover of those groups. Similar effects have been argued for diversification in metacommunities of ecologically neutral sexual species (Barraclough 2010).

Multi-locus sequence data for bacteria provide suitable data for testing for the presence of non-inter-recombining groups. For example, Didelot et al. (2011) assembled a dataset sequencing 10 % of the genome from each of 114 isolates of *Salmonella enterica*. They used population genetic analyses implemented in the software ClonalFrame and Structure to identify population structure within the clade caused by reduced recombination between particular sub-lineages. Five sub-clades were identified by both approaches that contained groups of isolates exchanging genes by recombination within groups but at a lower rate between groups. They called these groups potential ‘incipient species’. The overall picture of diversity was more complex, however, as many isolates failed to fall within one of these distinct groupings and instead shared genes more widely (including taking up genes from one of the five sub-clades). This pattern might reflect a tendency for bacterial diversity not to fall exclusively into readily distinguishable units based on barriers to gene exchange, or it might result from recent and ongoing diversification in this clade (which can yield similarly complex patterns of ancestry in sexual species complexes via incomplete lineage sorting).

Balbi et al. (in preparation) have developed an alternative approach to identifying non-recombining groups of bacteria. They extended the infinite alleles model of Fraser et al. (2005) to optimize clades within which there is recombination but between which there is none, and implemented it on multi-locus sequence type data for house-keeping genes from environmental samples of the *Bacillus cereus* complex. The method revealed five species-like clusters co-occurring at fine-scales within a 100 m square plot of grassland. The same groupings were recovered in a wider analysis including all isolates from the online database of *Bacillus cereus* multi-locus sequence type data (see also Raymond et al. 2010).

Other mechanisms for preventing gene exchange between diverging populations have been identified, in addition to the decay of homologous recombination with increasing genetic distance. For example, Carrolo et al. (2009) showed how changes in the control of competence in *Streptococcus pneumoniae*—the state in which cells are able to take up and integrate DNA into their genome by transformation (i.e. homologous recombination)—are associated with two co-occurring, but non-inter-recombining groups.

Competence in *S. pneumoniae* is controlled by a 17- amino acid quorum-sensing pheromone called competence-stimulating peptide (CSP). Accumulation of CSP in the medium induces competence, which includes not only the uptake of DNA but also the capacity to lyse non-competent cells, which enhances the release of DNA fragments to be taken up by competent cells. Several distinct CSPs have been identified, but most strains produce one of two varieties, CSP-1 and CSP-2. Cells with one variety are unable to respond to the other variety of signaling peptide. By comparing the distribution of phenotypes with inferences of gene exchange from multi-locus sequence data, the existence of two genetically isolated populations defined by phenotypes was demonstrated. Moreover, the spread of antibiotic resistance mutations was constrained by these boundaries: resistance to most antibiotics was restricted to a particular phenotype.

Similar mechanisms might operate to control the transfer of genes by phage-mediated transduction, if isolates varied in their susceptibility to phage infection. When gene exchange is under the control of such mechanisms (in contrast to simply decay in homologous recombination rates with genetic divergence), it is possible that selection could strengthen species boundaries. For example, if recombinant genotypes between two divergent strains suffered a fitness disadvantage, selection could promote divergence of control mechanisms to reduce gene exchange in a manner equivalent to reinforcement in sexual organisms (Coyne and Orr 1997). Alternatively, direct selection on, for example, signaling molecules to increase efficiency of communication in different environments might lead as an incidental by-product to barriers to recombination. The latter corresponds to the evolution of reproductive isolation in sexual organisms as a pleiotropic consequence of selection on ecological traits (Sobel et al. 2010).

One area of opportunity for further understanding the role of barriers to recombination in bacterial diversification is experimental evolution. Bacteria have long been the study organisms of choice for evolution experiments in the laboratory (Buckling et al. 2009), including studies of adaptive divergence into ecotypes (Kassen et al. 2004), but surprisingly few experiments have studied the evolution of barriers to recombination. Vulic et al. (1999) compared patterns of recombination between replicate lines of *Escherichia coli* evolved for 20,000 generations and showed how incipient genetic barriers could evolve. Future studies could use similar approaches to explore the drivers of barriers to gene exchange. For example, the evolution of barriers to gene exchange could be compared in lines experiencing parallel versus divergent selection pressures. Such experiments could be repeated with study strains with different mechanisms of gene exchange. It might not prove possible to simulate the entire process of speciation in the laboratory, even for prokaryotes with fast generation times.

However, the effects of selection and origin of reduced recombination between diverging populations can be studied in greater depth, and therefore provide evidence for mechanisms invoked by observational studies using retrospective sequence analysis of bacterial clusters.

To conclude, similarly to the field of eukaryote speciation, there remain differing opinions on how important barriers to recombination are for prokaryote speciation. Some argue that recombination is never strong enough to oppose speciation driven by divergent selection (Wiedenbeck and Cohan 2011). For others, the development of barriers to recombination still plays an important role in allowing divergence to occur, even if (as seems likely) divergence is predominantly driven by selection (e.g. Sheppard et al. 2008). The examples described above show how the importance of recombinational barriers for divergence can be investigated: more studies are now needed that consider these mechanisms in tandem with ecological mechanisms, discussed further in the next section, across a range of taxa with differing levels and mechanisms of gene exchange.

Ecological Speciation

Frederick Cohan and colleagues pioneered evolutionary definitions of bacterial species (and of clonally reproducing organisms in general) by considering how specialization to different ecological niches causes diversification (Cohan 2001, 2006). Even with strictly clonal reproduction, in which traditional ideas of speciation based on reproductive isolation do not apply, specialization to distinct niches can cause the emergence of independently evolving genetic clusters. Mutations that are beneficial for survival and reproduction in a particular niche spread within the genetic cluster of individuals adapted to that niche but not into genetic clusters adapted to other niches. Variation within each cluster depends on the effective population size and the frequency with which selective sweeps occur within each cluster. Genetic divergence between clusters increases linearly with time, as long as the clusters remain independently limited, i.e. as long as descendants of one cluster cannot replace the descendants of the other cluster. Therefore, as long as distinct ecological niches are present, and those niches persist long enough for genetic divergence to occur, then discrete and independently evolving clusters are expected to evolve. In some scenarios, the pattern of selection pressures across a clade could lead to more complex scenarios, for example with globally beneficial mutations occasionally spreading and reducing variation in the clade, followed by specialization and divergence into distinct locally adapted clusters. Or there could be hierarchical levels of divergent selection producing clustering at

multiple levels: for example, a set of species sharing the same abiotic environmental niche, but diversified into different resource use niches (see Wiedenbeck and Cohan 2011).

Koepfel et al. (2008) used simulations of these originally verbal models to test for the existence of discrete clusters (in their terminology, ecotypes) in members of the in *Bacillus simplex* and *Bacillus subtilis*–*Bacillus licheniformis* complex from dry canyons in Israel. Based on analyses of gene trees reconstructed from four genes, they found evidence for nine ecotypes in *B. simplex* and 17 ecotypes in the *B. subtilis*–*licheniformes* complex. Ecological distinctiveness was further evident from significant habitat differences and differences in heat tolerance between ecotypes. The simulation model assumes that ecological divergence is the force behind divergence into clusters. In principle alternative processes could lead to distinct clustering, such as geographical isolation (Barraclough et al. 2003; Fontaneto et al. 2007). Whether this is a common mechanism in bacteria is questionable, due to presumed high dispersal rates (Roberts and Cohan 1995; Whitaker 2006), although there is distance decay of genetic similarity in bacterial populations (Vos et al. 2009) and it remains theoretically possible. An alternative approach to detect clusters is therefore to remain agnostic about mechanism and to test for independently evolving genetic clusters. One such method is the generalized mixed Yule coalescent (GMYC) approach, which uses likelihood models of evolutionary branching to detect genetic clusters separated from each other by longer internal branches. First developed for asexual rotifers and for species delimitation in groups of insects (Pons et al. 2006; Fontaneto et al. 2007), it has also been applied to 16S data for bacteria (Barraclough et al. 2009). There are pros and cons to the two approaches: the Koepfel et al. (2008) simulations have the advantage of estimating parameters of biological interest, but as simulation methods they are more computer-intensive and cannot readily be applied to large datasets of the scale of Acinas et al. (2004) or Barraclough et al. (2009).

Other studies have tested ecotype theories of bacterial speciation by combining genetic and ecological data. Hunt et al. (2008) used associations between genetic clades and ecological characteristics of the size fraction of particles on which the bacteria grow (with free-living bacteria being sampled in the smallest size fraction) and seasonality to delimit ecotypes in bacterial plankton (family: Vibrionaceae). In principle, incorporation of ecological data could greatly enhance detection of species entities, compared to the use of arbitrary DNA markers (Barraclough 2010). However, for many bacteria, key ecological differences such as specialized resource use are likely to be hard to quantify relative to assembling DNA sequence data. Whole

genome sequencing offers major opportunities: with whole genomes sequenced from multiple isolates, population genetic tests can be used to identify regions of the genome under divergent selection (Vos 2011), and these can be used to test directly for the presence of ecotypes. For example, Luo et al. (2011) sequenced whole genomes of environmental isolates of *E. coli* and identified sets of genes associated with environmental versus enteric (i.e. living within host guts) lifestyles. Environmental isolates and enteric isolates differed in sets of genes relating to the use of particular energy sources, for example diol utilization in environmental isolates compared with fucose utilization in enteric isolates. Enteric isolates also contained several prophage genes in line with the distinct viral communities present in hosts compared to the environment. Such data provide direct evidence for ecotype divergence, yet are possible to compile across broad samples of isolates, necessary for delimiting units of diversification within wider clades.

Ecotype theories as originally proposed apply particularly to populations with low rates of recombination. If recombination rates were high enough, and if two ecotypes came into contact and exchanged genes, then genes associated with the two ecological niches would be shuffled up (Felsenstein 1981). In this case, divergence and coexistence in sympatry would require an additional mechanism to prevent gene exchange between the two emerging ecotypes—equivalent to the need for traits maintaining reproductive isolation in sexual eukaryotes. Alternatively, if only one or a few gene regions were important for adapting to the two niches, then clustering might arise just at those regions—arguably equivalent to adaptive polymorphism or ecotypes in sexual species if just based on one genomic region. Under these scenarios, marker genes such as 16S rRNA or the multiple housekeeping genes analyzed for multi-locus approaches would only show genetic clustering if they were linked to ecologically divergent genes or if divergent selection led to mechanisms for restricting gene exchange between ecotypes across the entire genome.

A recent study of the Archaeon *Sulfolobus islandicus* considered the possibility of genomic ‘islands’ of divergence in detail (Cadillo-Quiroz et al. 2012). By sequencing whole genomes from strains across several hot springs, two coexisting groups were identified that had lower levels of homologous recombination between them than within them. However, based on patterns of recombination and divergence across the genome, the authors inferred that there was no decay in recombination rates purely with genetic distance (as discussed above). Instead, they argued that ecological divergence was driving and maintaining genetic divergence, which varied in extent across the genome and was most pronounced in three genomic islands containing genes of putative ecological significance

(Cadillo-Quiroz et al. 2012). Although the reduced recombination between the two forms was not unambiguously attributed to ecological forces (there remained the possibility for additional undetected mechanisms), the pattern of genomic divergence matched theoretical expectations under ecological divergence.

Whether recombination would act against ecological divergence depends on the relative strength of selection and the rate of recombination. Cohan (1994) and Wiedenbeck and Cohan (2011) argue that recombination rates in bacteria are never high enough to counteract adaptive divergence between populations. However, as discussed in the BARRIERS TO RECOMBINATION section, recombination rates are high enough in many bacteria to maintain cohesion in neutral markers (Fraser et al. 2007); for example, marker loci were found to be at linkage equilibrium in a study of recombination and population structure in *Streptomyces flavogriseus* (Doroghazi and Buckley 2010). Furthermore, the example of phenotypes in *S. pneumoniae* demonstrates that ecologically divergent taxa are sometimes associated with specific mechanisms that prevent gene exchange (Carrolo et al. 2009). Recent contact between ecologically divergent populations has also been proposed to lead to loss of diversity through genetic exchange (Sheppard et al. 2008). In the light of these examples, and because recombination and divergent selection are the two key processes expected in theory to shape patterns of diversification, it remains important to consider both processes in understanding patterns and causes of bacterial diversification.

Horizontal Gene Transfer

The ideas outlined in the “Barriers to recombination” section concern recombination occurring between genetically similar isolates, which is broadly analogous to recombination in sexual populations. In prokaryotes, more frequently than in multicellular eukaryotes, gene exchange between distant relatives also occurs (Ochman et al. 2000). Horizontal gene transfer (HGT) might occur initially via a vector such as plasmids or phages, and genes are later incorporated into the host cell genome. Transformation might also play a role, although this process usually requires regions of homology between the donor and recipient, and is therefore less likely between distant relatives. Gene exchange between distant relatives, almost by definition, occurs too rarely to act as a homogenizing force on core genomes. Instead, it can influence patterns of diversity by transferring adaptations between species. There are several possible outcomes (discussed in Wiedenbeck and Cohan 2011):

- (1) Horizontal transfer might enhance the recipient’s fitness in its own niche. This could be an important mechanism for bacterial adaptation, but it is not expected to impact on patterns of diversity, as donor and recipient would still occupy the same niches as before transfer occurred.
- (2) Horizontal transfer might create a genotype able to occupy a new niche not accessible to either the donor or recipient. If there were no cost to fitness in the ancestral niche, then this would lead to niche expansion but not divergence into two distinct species (Wiedenbeck and Cohan 2011). Divergence additionally requires that there is a trade-off between fitness in the new niche and the ancestral niches, in which case the new recombinant genotype will diverge into a new and distinct genetic cluster within the clade. This mechanism is akin to hybrid speciation in sexual eukaryotes (Arnold and Martin 2010).
- (3) HGT might transfer a niche-partitioning trait involved in coexistence of species within a clade, i.e. genes enabling a species to occupy a given niche are transferred into a second species occupying a different niche. If the recombinant genotype had higher fitness than the donor genotype in the donor’s niche, the recombinant genotype would displace the donor in its own niche. This would lead to a short-term decline in diversity, as the donor is driven extinct. If, however, there remained a fitness trade-off between the two niches, the recombinant genotype would start to diverge from the population occupying the recipient’s original niche, restoring diversity. Alternatively, the recombinant genotype might have lower fitness than either the original recipient or donor genotypes. If adaptation to a particular niche involves multiple genes dispersed through the genome, then a single event is unlikely to transfer sufficient material to convey enough fitness benefit in the donor’s niche. Even if a niche trait can be encoded in a single piece of DNA, the gene might have negative interactions with the recipient genome, because it has not evolved within that genetic background. In this case, the recombinant genotype would be selected against and have no impact on diversity patterns in the clade.
- (4) Horizontal transfer of a gene encoding for barriers to gene exchange, such as the phenotypes identified in *S. pneumoniae* (Carrolo et al. 2009), could remove barriers to recombination between species. With sufficiently high levels of gene exchange for a long enough period, this could erode genetic differences between them, although any gene regions under strong enough divergent selection would be preserved as divergent between the species—genes

under uniform directional selection in both species would spread rapidly between them.

In general, HGT provides an additional source of variation for differences between species. The same kinds of variation could also arise, in principle, by mutation or rearrangements within a single genome, but *de novo* mutations are unlikely to convey as extreme changes in function as the acquisition of novel, pre-adapted genes (Perron et al. 2011). Certain traits are more likely to be associated with HGT than others (Wiedenbeck and Cohan 2011). Many known cases of HGT relate to environmental-filtering (also called beta-niche traits in the ecological literature, Ackerly 2003), i.e. the acquisition of a trait like antibiotic resistance or heavy metal tolerance without which survival in a given habitat is impossible. In these cases, there is a strong selective filter on bacteria invading the habitat to have that trait, and such habitats therefore select for either bacteria already possessing the trait or those able to acquire it by HGT: *de novo* evolution of the trait within a single genome is less likely. (Although note that the fitness benefits of genes transferred on plasmids can also be frequency-dependent, meaning that not all inhabitants require the trait, Ellis et al. 2007). Perhaps these are also traits that can be endowed by a single gene or relatively few genes. In contrast, niche-partitioning traits that differ among co-occurring species in a community (also called alpha-niche traits in the ecological literature, Ackerly 2003) might more often depend on resource utilization profiles and the relative expression of suites of metabolic pathways that are less easily transferred in a single event. Furthermore, there is not the same selective filter for bacteria to acquire alpha niche traits as there is for beta niche traits. Environments requiring traits involving point substitutions across multiple loci might select for invasion by mutator genotypes, rather than those able to acquire genes by horizontal transfer.

As argued by Wiedenbeck and Cohan (2011), comparisons of closely related strains are needed to test the role of HGT in bacterial speciation. Many studies have inferred HGT by comparing more distantly related taxa, which confirms the importance of HGT for acquisition of new functions, but does not necessarily demonstrate that HGT played a major role in the initial ecological divergence or emergence of barriers to recombination leading to divergence of a common ancestor. Cases of HGT have been shown to distinguish closely related ecotypes found in the enteric versus environmental *E. coli* discussed above (Luo et al. 2011), in gut versus dairy ecotypes of *Lactobacillus* (O'Sullivan et al. 2009), in isolates of *Pseudomonas putida* from polluted versus unpolluted soil (Wu et al. 2011), and marine bacterial plankton from high and low phosphate regions (Martiny et al. 2009, reviewed in Wiedenbeck and Cohan 2011).

Much of what we know about the evolutionary consequences of HGT comes from retrospective sequence analyses. As more and more genome sequence data accumulate, there is scope for more precise mapping of arenas of HGT. Do successful HGT events purely reflect opportunity or are there mechanisms that restrict transfers to subsets of taxa within a wider potential network of donors and recipients? Metagenome sequencing offers potential for reconstructing networks of HGT within communities of co-occurring species (c.f. Smillie et al. 2011), perhaps combined with experimental manipulations perturbing a community, for example by adding an antibiotic and tracking origins of resistance by *de novo* mutation versus transfer of pre-existing resistance mutations. There is also a need for experimental studies investigating the costs and benefits of HGT. Most laboratory studies of evolution in bacteria to date have used clonal populations (but see Cooper 2007; Baltrus et al. 2008; Perron et al. 2011), and there is a need to establish a broader set of study systems across the spectrum of levels of recombination and with different mechanisms of transferring genes.

Integrating Ecological and Recombinational Approaches

Many biologists remain skeptical about using the term 'species' for bacteria, and qualify its use with 'whatever they are' or the assertion that bacteria do not diversify into species. However, application of evolutionary concepts to bacteria as described above shows that similar processes to those operating in sexual eukaryotes can lead to similar patterns of phenotypic and genetic clustering, and to the emergence of non-recombining groups, as observed in sexual clades. Patterns of speciation are often complex in sexual eukaryotes as well, and different signatures of independent evolution can give different answers—some of the uncertainty regarding species limits is shared in both kinds of organisms. The wide variety of mechanisms for recombination (or its absence) in bacteria does lead to a richer range of possible outcomes than in sexual eukaryotes, but population genetic theory and analyses provide the tools to navigate and understand this range of outcomes.

More quantitative investigations are needed into patterns of diversity in representative bacterial clades and the evolutionary forces shaping those patterns. Do study clades fall into simple, independently evolving units of diversity equivalent to species, or is a more complex model of diversity needed encompassing different hierarchical levels of interactions? Do the patterns we observe across clades with different levels of recombination match those

predicted from theory? For example, with samples of multiple genes from large numbers of isolates across a wider clade, it is possible to test directly for patterns of recombination among isolates: can we detect discrete groups that represent arenas of recombination? Recent studies have begun such comparisons (e.g. Didelot et al. 2011, Cadillo-Quiroz et al. 2012), although the key step of delimiting groups (i.e. deciding which grouping to assign as evolutionary significant units) mostly still involves informal approaches to choose which groupings to consider. To answer the questions we pose, formal statistical methods are needed that consider alternative hypotheses for grouping of isolates against a null model that all isolates belong to a single clade with a specified homogeneous recombination rate. The Structure software performs this kind of assignment, but assumes that populations are unrelated, rather than derived by evolutionary branching from a common ancestor. Ideally, methods that combine population assignment with genealogical analysis of evolutionary history are needed.

By sampling ecologically relevant genes, such as those involved in virulence or resource use, we can similarly test directly for units of divergent selection: groups of isolates that experience purifying selection on particular protein sequences but for which there is divergent selection acting between those groups. Current methods for identifying positive selection do so either without considering species boundaries or using a priori assignment of species boundaries (as in the test of McDonald and Kreitman 1991). New statistical methods are needed that use patterns of genetic divergence to test alternative assignments for the units of diversity being targeted by divergent selection (c.f. Fontaneto et al. 2007). For example, does divergent selection on genes involved in virulence of pathogenic strains act on the same units of diversity identified from analyses of barriers to recombination from house-keeping genes used in standard MLST approaches? The same analyses can be performed for genes on plasmids and other vectors associating with host genomes to determine arenas for plasmid transfer and HGT.

The findings from alternative tests can then be synthesized to determine the mode of diversification in the clade. Do the units of diversity based on patterns of divergent selection on ecologically important genes coincide with those identified based on patterns of recombination? One possible outcome is that all of these tests identify the same arenas of evolutionary interactions, and that the clade has indeed diversified into discrete entities that we can call species (Fig. 1a). An alternative is that the core genome displays consistent units but some genes are exchanged and interact within wider arenas (Fig. 1b). It might still be useful then to use the term species for the patterns obtained from the core genome, but to delimit and identify

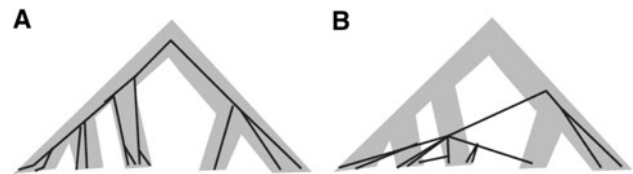


Fig. 1 Two alternative scenarios for diversity patterns in a hypothetical clade of bacteria. **a** Simple case in which patterns of ancestry and genetic clustering of core genes (consensus species tree in *grey shading*) and ecological genes (one shown in *black*) coincide. **b** Complex scenario in which ecological genes interact and recombine across a broader set of individuals than the core genes, therefore showing clusters that differ from those of core genes. In case **a** analysis of core house-keeping genes (summarized by *grey branches*) indicates five clusters, and the ancestry of the ecological gene fits with this scenario except for stochastic differences due to incomplete lineage sorting (which could be distinguished from hard incongruence using statistical genealogical analyses). In case **b**, however, the pattern of diversity of the ecological gene differs significantly from the ancestry of the core genes: instead of five separate clusters, there are two, with four of the *grey* clusters sharing copies of the ecological gene derived from a more recent common ancestor than expected given the branching pattern of the core genes

additional arenas of interaction (e.g. via horizontal gene transfer) at higher levels. Or there might be such a discrepancy of histories and interactions across different parts of the genome that the species-concept can be abandoned altogether in favour of a more complex model of diversity that requires more complex representation. In such cases, the search would shift to understanding how genes with disparate histories can come together to form functioning genomes. Given the likely ubiquity of epistatic interactions among genes (de Visser et al. 2011), it seems unlikely that this extreme scenario can apply except among bacteria with very similar genomes (but see Lawrence and Retchless 2010). The mechanisms inferred from observational studies can also be tested with laboratory evolution experiments using culturable model taxa. Many evolution experiments have treated bacteria as tractable model systems with which to test ideas for multicellular eukaryotes: to understand mechanisms of bacterial evolution in their own right, a broader range of experimental organisms is needed (Buckling et al. 2009), especially using strains with a wider range of mechanisms of gene exchange rather than just clonally reproducing laboratory strains.

Recent years have seen a revolution in the availability of DNA sequence data and in the development of tools to extract information from those data, such as identifying genes under selection within whole genomes and reconstructing recombination events. Tests of the forces generating diversity and of the evolutionary nature of units of diversity remain rarer. Yet, identifying the nature of arenas of selection and recombination in bacteria is critical to predicting how bacteria evolve in response to changing environments: how far will beneficial mutations spread and

what will the consequences be for functional diversity (i.e. which other lineages will be driven extinct by the spread of new forms)? The answers to such questions rely on identifying the nature of species boundaries. The concepts in this paper are not new, but by revisiting ideas and outlining testable predictions from sequence data, we hope to stimulate quantitative tests of alternative models of bacteria diversification.

Acknowledgments These ideas were developed on BBSRC grant BB/G004250/1. We thank Albert Phillimore and two anonymous reviewers for discussions and Frederick Cohan for his many contributions on bacterial speciation, which underpin most of the ideas in this manuscript.

References

- Achtman, M., & Wagner, M. (2008). Microbial diversity and the genetic nature of microbial species. *Nature Reviews Microbiology*, 6, 431–440.
- Acinas, S. G., Klepac-Ceraj, V., Hunt, D. E., Pharino, C., Ceraj, I., Distel, D. L., et al. (2004). Fine-scale phylogenetic architecture of a complex bacterial community. *Nature*, 430, 551–554.
- Ackerly, D. D. (2003). Community assembly, niche conservatism, and adaptive evolution in changing environments. *International Journal of Plant Sciences*, 164, S165–S184.
- Arnold, M. L., & Martin, N. H. (2010). Hybrid fitness across time and habitats. *Trends in Ecology & Evolution*, 25, 530–536.
- Balbi, K. J., Barraclough, T. G., & Ellis, R. J. A population genetic approach to defining bacterial species: Coexistence of multiple, independently-recombining bacterial lineages (in preparation).
- Baltrus, D. A., Guillemin, K., & Phillips, P. C. (2008). Natural transformation increases the rate of adaptation in the human pathogen *helicobacter pylori*. *Evolution*, 62, 39–49.
- Barraclough, T. G. (2010). Evolving entities: Towards a unified framework for understanding diversity at the species and higher levels. *Philosophical Transactions of the Royal Society of London. Series B*, 365, 1801–1813.
- Barraclough, T. G., Birky, C. W., & Burt, A. (2003). Diversification in sexual and asexual organisms. *Evolution*, 57, 2166–2172.
- Barraclough, T. G., Hughes, M., Ashford-Hodges, N., & Fujisawa, T. (2009). Inferring evolutionarily significant units of bacterial diversity from broad environmental surveys of single-locus data. *Biology Letters*, 5, 425–428.
- Buckling, A., Maclean, R. C., Brockhurst, M. A., & Colegrave, N. (2009). The Beagle in a bottle. *Nature*, 457, 824–829.
- Cadillo-Quiroz, H., Didelot, X., Held, N. L., Herrera, A., Darling, A., et al. (2012). Patterns of gene flow define species of thermophilic Archaea. *PLoS Biology*, 10(2), e1001265.
- Carrolo, M., Pinto, F. R., Melo-Cristino, J., & Ramirez, M. (2009). Phenotypes are driving genetic differentiation within *Streptococcus pneumoniae*. *BMC Microbiology*, 9, 191. doi:10.1186/1471-2180-9-191.
- Cohan, F. M. (1994). The effects of rare but promiscuous genetic exchange on evolutionary divergence in prokaryotes. *American Naturalist*, 143, 965–986.
- Cohan, F. M. (2001). Bacterial species and speciation. *System Biology*, 50, 513–524.
- Cohan, F. M. (2006). Towards a conceptual and operational union of bacterial systematics, ecology, and evolution. *Philosophical Transaction of the Royal Society of London. Series B*, 361, 1985–1996.
- Cooper, T. F. (2007). Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biology*, 5, 1899–1905.
- Coyne, J. A., & Orr, H. A. (1997). Patterns of speciation in *Drosophila* revisited. *Evolution*, 51, 295–303.
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sunderland, MA: Sinauer Associates.
- Curtis, T. P., Head, I. M., Lunn, M., Woodcock, S., Schloss, P. D., & Sloan, W. T. (2006). What is the extent of prokaryotic diversity? *Philosophical Transaction of the Royal Society of London. Series B*, 361, 2023–2037.
- de Visser, J. A. G. M., Cooper, T. F., & Elena, S. F. (2011). The causes of epistasis. *Philosophical Transactions of the Royal Society of London. Series B*, 278, 3617–3624.
- Didelot, X., Bowden, R., Street, T., Golubchik, T., Spencer, C., McVean, G., et al. (2011). Recombination and Population Structure in *Salmonella enterica*. *PLoS Genetics*, 7(7), e1002191. doi:10.1371/journal.pgen.1002191.
- Doolittle, W. F., & Zhaxybayeva, O. (2009). On the origin of prokaryotic species. 2009. *Genome Research*, 19, 744–756.
- Doroghazi, J. R., & Buckley, D. H. (2010). Widespread homologous recombination within and between *Streptomyces* species. *ISME Journal*, 4, 1136–1143.
- Eisen, J. A. (2007). Environmental shotgun sequencing: Its potential and challenges for studying the hidden world of microbes. *PLoS Biology*, 5, 384–388.
- Ellis, R. J., Lilley, A. K., Lacey, S. J., Murrell, D., & Godfray, H. C. J. (2007). Frequency-dependent advantages of plasmid carriage by *Pseudomonas* in homogeneous and spatially structured environments. *ISME Journal*, 1, 92–95.
- Felsenstein, J. (1981). Skepticism towards Santa Rosalia, or why are there so few kinds of animals? *Evolution*, 35, 124–138.
- Fisher, R. A. (1930). *The genetical theory of natural selection*. Oxford: Oxford University Press.
- Fontaneto, D., Herniou, E. A., Boschetti, C., Caprioli, M., Melone, G., Ricci, C., et al. (2007). Independently evolving species in asexual bdelloid rotifers. *PLoS Biology*, 5, 914–921.
- Fraser, C., Hanage, W. P., & Spratt, B. G. (2005). Neutral microepidemic evolution of bacterial pathogens. *Proceedings of the National Academy of Sciences of the United States of America*, 102, 1968–1973.
- Fraser, C., Hanage, W. P., & Spratt, B. G. (2007). Recombination and the nature of bacterial speciation. *Science*, 315, 476–480.
- Gevers, D., Cohan, F. M., Lawrence, J. G., Spratt, B. G., Coenye, T., Feil, E. J., et al. (2005). Re-evaluating prokaryotic species. *Nature Reviews Microbiology*, 3, 733–739.
- Hanage, W. P., Fraser, C., & Spratt, B. G. (2006). Sequences, sequence clusters and bacterial species. *Philosophical Transaction of the Royal Society of London. Series B*, 361, 1917–1927.
- Hunt, D. E., David, L. A., Gevers, D., Preheim, S. P., Alm, E. J., & Polz, M. F. (2008). Resource partitioning and sympatric differentiation among closely related bacterioplankton. *Science*, 320, 1081–1085.
- Kassen, R., Llewellyn, M., & Rainey, P. B. (2004). Ecological constraints on diversification in a model adaptive radiation. *Nature*, 431, 984–988.
- Keymer, D. P., & Boehm, A. B. (2011). Recombination shapes the structure of an environmental vibrio cholerae population. *Applied and Environmental Microbiology*, 77, 537–544.
- Koepfel, A., Perry, E. B., Sikorski, J., Krizanc, D., Warner, A., Ward, D. M., et al. (2008). Identifying the fundamental units of bacterial diversity: A paradigm shift to incorporate ecology into bacterial systematics. *Proceedings of the National Academy of Sciences of the United States of America*, 105, 2504–2509.
- Lawrence, J. G., & Retchless, A. C. (2010). The myth of bacterial species and speciation. *Biology and Philosophy*, 25, 569–588.

- Lorenz, M. G., & Wackernagel, W. (1994). Bacterial gene-transfer by natural genetic-transformation in the environment. *Nature Reviews Microbiology*, *58*, 563–602.
- Luo, C., Walk, S. T., Gordon, D. M., Feldgarden, M., Tiedje, J. M., & Konstantinidis, K. T. (2011). Genome sequencing of environmental *Escherichia coli* expands understanding of the ecology and speciation of the model bacterial species. *Proceedings of the National Academy of Sciences of the United States of America*, *108*, 7200–7205.
- Martiny, A. C., Huang, Y., & Li, W. Z. (2009). Occurrence of phosphate acquisition genes in *Prochlorococcus* cells from different ocean regions. *Environmental Microbiology*, *11*, 1340–1347.
- McDonald, J. H., & Kreitman, M. (1991). Adaptive protein evolution at the *adh* locus in *Drosophila*. *Nature*, *351*, 652–654.
- Norman, A., Hansen, L. H., & Sorensen, S. J. (2009). Conjugative plasmids: vessels of the communal gene pool. *Philosophical Transactions of the Royal Society of London. Series B*, *364*, 2275–2289.
- Nosil, P., Funk, D. J., & Ortiz-Barrientos, D. (2009). Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, *18*, 375–402.
- O'Sullivan, O., O'Callaghan, J., Sangrador-Vegas, A., McAuliffe, O., Slattery, L., Kaleta, P., et al. (2009). Comparative genomics of lactic acid bacteria reveals a niche-specific gene set. *BMC Microbiology*, *9*, 50. doi:10.1186/1471-2180-9-50.
- Ochman, H., Lawrence, J. G., & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature*, *405*, 299–304.
- Ochman, H., Lerat, E., & Daubin, V. (2005). Examining bacterial species under the specter of gene transfer and exchange. *Proceedings of the National Academy of Sciences of the United States of America*, *102*, 6595–6599.
- Papke, R. T., Zhaxybayeva, O., Feil, E. J., Sommerfeld, K., Muise, D., & Doolittle, W. F. (2007). Searching for species in haloarchaea. *Proceedings of the National Academy of Sciences of the United States of America*, *104*, 14092–14097.
- Perron, G. G., Lee, A. E. G., Wang, Y., Huang, W. E., & Barraclough, T. G. (2011). Bacterial recombination promotes the evolution of multi-drug-resistance in functionally diverse populations. *Philosophical Transactions of the Royal Society of London. Series B*. (Online early).
- Pons, J., Barraclough, T. G., Gomez-Zurita, J., Cardoso, A., Duran, D. P., Hazell, S., et al. (2006). Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *System Biology*, *55*, 595–609.
- Qin, J. J., Li, R. Q., Raes, J., Arumugam, M., Burgdorf, K. S., Manichanh, C., et al. (2010). A human gut microbial gene catalogue established by metagenomic sequencing. *Nature*, *464*, 59–65.
- Rappe, M. S., & Giovannoni, S. J. (2003). The uncultured microbial majority. *Annual Review of Microbiology*, *57*, 369–394.
- Raymond, B., Wyres, K. L., Sheppard, S. K., Ellis, R. J., & Bonsall, M. B. (2010). Environmental factors determining the epidemiology and population genetic structure of the *Bacillus cereus* group in the field. *PLoS Pathogens*, *6*(5), e1000905. doi:10.1371/journal.ppat.1000905.
- Roberts, M. S., & Cohan, F. M. (1993). The effect of DNA-Sequence divergence on sexual isolation in *Bacillus*. *Genetics*, *134*, 402–408.
- Roberts, M. S., & Cohan, F. M. (1995). Recombination and migration rates in natural populations of *Bacillus subtilis* and *Bacillus mojavensis*. *Evolution*, *49*, 1081–1094.
- Schloss, P. D., & Handelsman, J. (2005). Introducing DOTUR, a computer program for defining operational taxonomic units and estimating species richness. *Applied and Environmental Microbiology*, *71*, 1501–1506.
- Sheppard, S. K., McCarthy, N. D., Falush, D., & Maiden, M. C. J. (2008). Convergence of *Campylobacter* species: Implications for bacterial evolution. *Science*, *320*, 237–239.
- Smillie, C. S., Smith, M. B., Friedman, J., Cordero, O. X., David, L. A., & Alm, E. J. (2011). Ecology drives a global network of gene exchange connecting the human microbiome. *Nature* (Online early).
- Sobel, J. M., Chen, G. F., Watt, L. R., & Schemske, D. W. (2010). The biology of speciation. *Evolution*, *64*, 295–315.
- Stackebrandt, E., & Ebers, J. (2006). Taxonomic parameters revisited: Tarnished gold standards. *Microbiol Today*, *33*, 152–155.
- Tautz, D., Arctander, P., Minelli, A., Thomas, R. H., & Vogler, A. P. (2003). A plea for DNA taxonomy. *Trends in Ecology & Evolution*, *18*, 70–74.
- Templeton, A. (1989). The meaning of species and speciation: a population genetics approach. In D. Otte & J. Endler (Eds.), *Speciation and its consequences*. Sunderland, MA: Sinauer Associates.
- Venter, J. C., Remington, K., Heidelberg, J. F., Halpern, A. L., Rusch, D., Eisen, J. A., et al. (2004). Environmental genome shotgun sequencing of the Sargasso Sea. *Science*, *304*, 66–74.
- Vos, M. (2009). Why do bacteria engage in homologous recombination? *Trends in Microbiology*, *17*, 226–232.
- Vos, M. (2011). A species concept for bacteria based on adaptive divergence. *Trends in Microbiology*, *19*, 1–7.
- Vos, M., Birkett, P. J., Birch, E., Griffiths, R. I., & Buckling, A. (2009). Local adaptation of bacteriophages to their bacterial hosts in soil. *Science*, *325*, 833–834.
- Vulic, M., Lenski, R. E., & Radman, M. (1999). Mutation, recombination, and incipient speciation of bacteria in the laboratory. *Proceedings of the National Academy of Sciences of the United States of America*, *96*, 7348–7351.
- Whitaker, R. J. (2006). Allopatric origins of microbial species. *Philosophical Transactions of the Royal Society of London. Series B*, *361*, 1975–1984.
- Whitaker, R. J., & Banfield, J. F. (2006). Population genomics in natural microbial communities. *Trends in Ecology & Evolution*, *21*, 508–516.
- Wiedenbeck, J., & Cohan, F. M. (2011). Origins of bacterial diversity through horizontal genetic transfer and adaptation to new ecological niches. *FEMS Microbiology Reviews*, *35*, 957–976.
- Wu, X., Monchy, S., Taghavi, S., Zhu, W., Ramos, J., & van der Lelie, D. (2011). Comparative genomics and functional analysis of niche-specific adaptation in *Pseudomonas putida*. *FEMS Microbiology Reviews*, *35*, 299–323.
- Zawadzki, P., Roberts, M. S., & Cohan, F. M. (1995). The log-linear relationship between sexual isolation and sequence divergence in *Bacillus* transformation is robust. *Genetics*, *140*, 917–932.